

# Panel Discussion: Engagement, Trust and Intimacy: Are these the Essential Elements for a ‘Successful’ Interaction Between a Human and a Robot?

Mari Velonaki, David Rye and Steve Scheduling

Centre for Social Robotics  
Australian Centre for Field Robotics J04  
The University of Sydney 2006, Australia  
{m.velonaki | d.rye | s.scheduling}@acfr.usyd.edu.au

## Introduction

As the capability of robots increases, and interactions between humans and machines become more complex, it is important for researchers to consider the potential for an emotional connection to be established between a human and a machine. The panel will discuss areas such as the development of engagement, trust and intimacy during human interaction with embodied robotic characters, and assigning ‘personalities’ and ‘emotional states’ to robotic characters that interact with people in public spaces.

It is intended that the panel will focus specifically on questions such as

- What is trust, and can you share it with a machine?
- What is intimacy, and can you share it with a machine?
- How do theories of behavior from different perspectives assist in understanding our complex relationships with robotic ‘others’?

One part of our collaborative work involves interactive robotics in a media arts environment. *Fish-Bird Circle B—Movement C* (figure 1) is an autokinetic artwork that aims to investigate the dialogical possibilities that exist between autokinetic objects (two robots disguised as wheelchairs) that have the ability to communicate with each other and with their audience through several modalities. The wheelchairs impersonate two characters (Fish and Bird) who fall in love but cannot be together due to ‘technical’ difficulties. In their shared isolation, Fish and Bird communicate intimately with one another via movement and text. Assisted by integrated printers, the wheelchairs write intimate letters to each other.

Spectators entering the installation space disturb the intimacy of the two objects, yet create the strong potential (or need) for other dialogues to exist. The spectator can see the traces of their previous conversation on the letters dropped to the floor, and may become aware of the disturbance that s/he has caused. Dialogue occurs kinetically through the wheelchairs’ ‘perception’ of the body language of the audience, and through the audience’s reaction to the ‘body language’ of the chairs. A common initial reaction to

the unexpected disturbance would be for Fish and Bird to converse about trivial subjects, like the weather... Through emerging dialogue, the wheelchairs may become more ‘comfortable’ with their observers, and start to reveal intimacies on the floor again.



Figure 1: *Fish-Bird Circle B—Movement C* (2005).

During many exhibitions of *Fish-Bird* we have observed participants who were clearly engaged with the artwork—and perhaps with the Fish and Bird characters—spending 30 minutes or more interacting with the robots, or returning to the exhibition day after day. We confidently believe that these repeated observations demonstrate engagement of the person with the (clearly non-anthropomorphic) machine. Does it, however, constitute some sort of relationship, albeit a lopsided one? Furthermore, it is clear that the human participant must trust the machine, at least to some extent: after all, they do share the same space.

These speculations, and others, lead us to propose the panel discussion, seeking to draw in experience and opinion from a variety of practitioners from disparate fields. How can the ‘relationship’ between a human and a machine be understood? Can it be predicted, and if so can

human response be shaped by the actions of a machine? Are the 'trust' and 'intimacy' only illusory and, if so, can the illusion be persuasive enough to become reality for the human?

The following section gives short biographies of the three invited panelists: Lola Cañamero, Karl F. MacDorman and Nishio Suichi. These are followed by brief overviews of the talks prepared by the three speakers. Unfortunately, Dr Cañamero's overview was not available at time of press. Conclusions shall have to wait until the panel discussion.

## Panelists

### Lola Cañamero

Lola Cañamero is Reader in Adaptive Systems in the School of Computer Science at the University of Hertfordshire in the UK, where she has worked since 2001. She received a BA and MA in Philosophy from the Complutense University of Madrid, Spain, and a PhD in Computer Science (1995) from the University of Paris-XI in France. Before her current position, she worked as a postdoctoral associate with Rodney Brooks at the MIT AI-Lab in the US, and with Luc Steels at the VUB AI-Lab in Belgium, and as a senior researcher at the Spanish Scientific Research Council, IIIA-CSIC. Her research lies in the areas of modeling emotions and their expression in autonomous and social agents (both robotic and synthetic), adaptive behavior, and human-robot interaction. She investigates these issues taking a biologically-inspired approach and from different perspectives: ethological, developmental, and evolutionary. She has authored or co-authored over 70 scientific papers in journals, refereed conferences and books. Dr Cañamero edited, with others, *Socially Intelligent Agents: Creating Relationships with Computers and Robots* (Kluwer 2002), and the special issue 'Achieving human-like qualities in interactive virtual and physical humanoids' of *I. J. of Humanoid Robotics* (2006). Since January 2004 she has coordinated the thematic area 'Emotion in Cognition and Action' of the EU-funded HUMAINE Network of Excellence on emotions in human-machine interaction, and since December 2006 she has lead the EU-funded Advanced Robotics project 'Felix Growing' which investigates socio-emotional development from an interdisciplinary perspective.

<http://homepages.feis.herts.ac.uk/~comqlc>

### Karl F. MacDorman

Karl F. MacDorman is Associate Professor of Informatics at Indiana University in the Human-Computer Interaction program. Dr MacDorman received his BA in computer science from University of California, Berkeley in 1988 and his PhD in machine learning and robotics from Cambridge University in 1996. Most recently MacDorman was an associate professor at Osaka University, Japan (2003–2005). Previously, he was assistant professor in the Department of Systems and Human Science at the same institution (1997–2000), and a supervisor (1991–1997) and research fellow (1997–1998) at Cambridge University. Dr MacDorman has also worked as a software engineer at Sun Microsystems and as chief technology officer for two venture companies. His research focuses on human-robot interaction and the symbol grounding problem. He co-organized the workshop 'Toward Social Mechanisms of Android Science' at CogSci 2005 and CogSci/ICCS 2006, the workshop 'Views of the Uncanny Valley' at IEEE Humanoids 2005, and the special session 'Psychological Benchmarks of Human-Robot Interaction' at IEEE Ro-Man 2006 and has edited special issues on these topics for *Connection Science* and *Interaction Studies*. He has published extensively in robotics, machine learning, and cognitive science.

<http://www.macdorman.com>

### Nishio Shuichi

Nishio Shuichi received the MSc degree in computer science from Kyoto University (1994) and joined Nippon Telegraph and Telephone Corporation (NTT) in 1994. He engaged in research on human emotion and in the development of commercial high-speed IPv6 networks, VoIP and multimedia applications (NTT-West). In 2005 Mr Nishio joined ATR Intelligent Robotics and Communication Laboratories in Kyoto, and is now working on robotic sensor networks, android science, and standardization of robotic technology. He is a Member of IEEE, IEICE and JSRE.

# Long-term Relationships as a Benchmark for Robot Personhood

Karl F. MacDorman<sup>1</sup> and Stephen J. Cowley<sup>2</sup>

<sup>1</sup>School of Informatics, Indiana University  
535 West Michigan Street  
Indianapolis IN 46202, USA

<sup>2</sup>School of Psychology  
University of Hertfordshire  
College Lane, Hatfield, Hertfordshire  
AL10 9AB, UK

## Extended Abstract

The human body constructs itself into a person by becoming attuned to the affective consequences of its actions in social relationships. Norms develop that ground perception and action, providing standards for appraising conduct. The body finds itself motivated to enact itself as a character in the drama of life, carving from its beliefs, intentions, and experiences a unique identity and perspective. If a biological body can construct itself into a person by exploiting social mechanisms, could an electromechanical body, a robot, do the same? To qualify for personhood, a robot body must be able to construct its own identity, to assume different roles, and to discriminate in forming friendships. Though all these conditions could be considered benchmarks of personhood, the most compelling benchmark, for which the above mentioned are prerequisites, is the ability to sustain long-term relationships. Long-term relationships demand that a robot continually recreate itself as it scripts its own future. Although personhood should not in principle be limited to one species, the most humanlike of robots will be best equipped for reciprocal relationships with human beings.

Robot personhood is a long-term goal of artificial intelligence. But we human beings are currently our only example of a species that can construct an identity from the social environment, align to inter-individual and socio-cultural norms, and coherently narrate motives and intentions with respect to those norms. And we tend to expect the same from each other. In designing a benchmark for robot personhood, therefore, the ability to form and maintain long-term relationships with people could be a useful measure.

The paper sets out the connection between the long-term goal of robot personhood and the benchmark of forming and maintaining a long-term relationship. Specifically, to develop even the beginnings of personhood, a machine

would have to act with sensitivity to socio-cultural norms to participate in a changing relationship. This view builds on two main sources: Robert Hinde and Daniel Dennett. Hinde shows us what it means to participate in a relationship and be transformed by it, and Dennett shows why this is important to our perceptions of human beings. As summarized by Ross and Dumouchel (2004, pp. 264–265), Dennett gives a useful starting point for simulations that begin with a third-person analysis of interaction:

Biological systems of the *H. sapiens* variety turn themselves into people—socially embedded teleological selves with narrated biographies in terms of these very beliefs and desires—by taking the intentional stance toward themselves. They can do this thanks to the existence, out in the environment, of public languages that anchor their interpretations to relatively consistent and socially enforced rules of continuity... [T]hey are incentivized to narrate themselves as coherent and relatively predictable characters, and to care deeply about the dramatic trajectories of these characters they become... [People] are partly constituted out of their social environments, both in the networks of expectations that give identity to them as people, and in the fact that the meanings of their own thoughts are substantially controlled by semantic systems that are collective rather than individual. They are thus not identical to their nervous systems, which are indeed constituted internally.

In the human case, we take the grounding of this transformation to lie in attachments: the surprisingly stable patterns of behavior that emerge between infant and caregiver towards the end of the first year of life (Ainsworth 1978; Bowlby 1969; Hinde 1987; Sullivan 1938). They are, moreover, the place in which children stabilize both those forms of sensorimotor activity that are themselves the rudiments of language and the kinds of rational choices

that enable parents to take an intentional stance to what they are doing. Later, of course, these same stable behavioral patterns will be the basis for children coming to use ‘words’ in taking this same stance towards their emergent selves.

Why do we take this view? First, it is problematic to define one of the goals of humanoid robotics as the creation of artificial human beings, because ‘human being’ is, at least partly, a biological category and ‘robot’ is an electromechanical category. This paradox can be sidestepped if we redefine the goal as the creation of an artificial person, while defining person with language that is free of speciesism (i.e., the presumption of human superiority; Regan 1983; Ryder 1989). This is suggested below:

BIOLOGICAL BODY →  
ATTACHED PROTO-PERSON → PERSON  
ELECTROMECHANICAL BODY →

There are advantages in regarding a person as an emergent entity. Instead of judging an entity’s personhood on the ground of whether it is conscious, we set the benchmark of whether it appears to be conscious. This is compatible with what would happen if an alien species of obvious intelligence arrived on earth. In determining whether to afford it rights, ethicists would focus on whether it were conscious and had feelings, thoughts, hopes, and fears—on whether it related to us in ways that resembled what might be expected of a person.

Nevertheless, to lose consciousness, as in a deep sleep, is not to lose personhood. Also, we consider many species of animals to be conscious but not necessarily to be persons.

And still, given the problem of “other minds,” we may have doubts about human simulacra, such as android robots (MacDorman and Ishiguro 2006a), ever being conscious, regardless of how humanlike their behavior might be. Since there is no consensus on why human beings are conscious, at this stage it is hard to predict whether machines could be conscious. However, if a robot looks and appears to act human, it may be hard to resist treating it as a fully conscious person (MacDorman and Ishiguro 2006b). Brief operational tests of intelligence, such as the Turing test (Turing 1950), in which a computer is expected to pretend to be human, are both too easy and too difficult. They are too easy, because a mindless program can fool ordinary people into thinking it is human (Block 1981; Harnad 1990; Weizenbaum 1966). They are too difficult, because a clever judge can devise questions that no computer however brilliant could answer as a human being would—namely, questions designed to tease apart its sub-cognitive architecture (French 1990). Clearly, the Turing test, whether conducted in its original form across a teleprinter or in its more recent robotic incarnations (Harnad 1992; 2000), suffers from speciesism. This may be one reason for the test’s waning significance (French 2000).

In focusing on whether machines can mimic persons, however, we write in the spirit of the Turing test. This depends, above all, in attempting to develop agents who relate to us in ways that are broadly true to life. The focus

on the extent to which robots can turn themselves into people enables us to bring to the fore—not the states that we reportedly have—but reporting on states. While it can be objected that this is oversimplified, we would be impressed by a machine that could react appropriately to our attitudes, hopes, expectations, and emotions. We would be even more impressed if, like a baby, it was able to use rudimentary language in reporting on some of the events of social life. In aiming to build robots that are capable of developing attachments, however, it is precisely this that comes to the fore.

## References

- Ainsworth, M.D.S., Biehat, M.C., Waters, E., and Wall, S. 1978. *Patterns of Attachment*. Hillsdale NJ: Erlbaum.
- Altmann, J. 1980. *Baboon Mothers and Infants*: Cambridge MA: Harvard University Press.
- Block, N. Psychologism and behaviorism. *Philosophical Review* 90: 5–43.
- Bowlby, J. 1969. *Attachment and Loss, Vol 1: Attachment*. London: Hogarth.
- Cowley, S. J. 2004. Contextualizing Bodies: How Human Responsiveness Constrains Distributed Cognition. *Language Sciences* 26: 565–591.
- Cowley, S. J. 2005. Beyond symbols: How Interaction Enslaves Distributed Cognition. In P. Thibault and C. Prevignano (eds), *Interaction Analysis and Language: Discussing the State-of-the-Art*, Forthcoming.
- Cowley, S. J. and K. MacDorman, K. F. 2007. What Baboons, Babies, and Tetris Players Tell Us About Interaction: A Biosocial View of Norm-based Social Learning. *Connection Science* 18: 363–378.
- French, R.M. 1990. Subcognition and the Limits of the Turing Test. *Mind* 99: 53–65.
- French, R.M. 2000. The Turing Test: The First Fifty Years. *Trends in Cognitive Sciences* 4: 115–121.
- Harnad, S. 1990. Lost in the Hermeneutic Hall of Mirrors: An Invited Commentary on Michael Dyer’s *Minds, Machines*, Searle and Harnad. *Journal of Experimental and Theoretical Artificial Intelligence* 2: 321–327.
- Harnad, S. 2000. Minds, Machines and Turing: The Indistinguishability of Indistinguishables. *Journal of Logic, Language, and Information* 9: 425–445.
- Harnad, S. 1992. The Turing Test is Not a Trick: Turing Indistinguishability is a Scientific Criterion. *SIGART Bulletin* 3: 9–10.
- Hinde, R. A. 1987. *Individuals, Relationships & Culture*. Cambridge: Cambridge University Press.
- MacDorman, K. F., and Ishiguro, H. 2006a. The Uncanny Advantage of Using Androids in Social and Cognitive Science Research. *Interaction Studies* 7: 297–337

- MacDorman, K. F., and Ishiguro, H. 2006b. Toward Social Mechanisms of Android Science: A CogSci2005 Workshop. *Interaction Studies* 7: 361–368.
- Regan, T. 1983. *The Case for Animal Rights*. Berkeley CA: University of California Press.
- Ross, D. 2004. Emotions as Strategic Signals. *Rationality and Society* 16: 251–286.
- Ryder, R. *Animal revolution: Changing attitudes towards speciesism*. Oxford: Blackwell, 1989.
- Sullivan, H. S. 1938. The Data of Psychiatry. *Psychiatry* 1: 121–134.
- Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59: 433–460.
- Weizenbaum, J. 1966. Eliza: A Computer Program for the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM* 9: 36–45.

# Artificial Humans for Understanding Human Presence

Hiroshi Ishiguro<sup>1</sup> and Shuichi Nishio<sup>2</sup>

<sup>1</sup> ATR Intelligent Robotics and Communication Laboratories and  
Department of Adaptive Machine Systems  
2-1 Yamada-oka Suita Osaka 565-0871, Japan  
ishiguro@ams.eng.osaka-u.ac.jp

<sup>2</sup>ATR Intelligent Robotics and Communication Laboratories  
2-2-2 Hikaridai Keihanna Science City Kyoto 619-0288, Japan  
nishio@ieee.org

## Extended Abstract

Why are people attracted to humanoid robots and androids? The answer is simple: because human beings are attuned to understand or interpret human expressions and behaviors, especially those that exist in their surroundings. As they grow, infants—who are supposedly born with the ability to discriminate various types of stimuli—gradually adapt and fine-tune their interpretations of detailed social clues from other voices, languages, facial expressions, or behaviors (Pascalis, Haan, and Nelson 2002). Perhaps due to this functionality of nature and nurture, people have a strong tendency to anthropomorphize nearly everything they encounter. This is also true for computers or robots. Recently, researchers' interests in robotics are shifting from traditional studies on navigation and manipulation to human-robot interaction. A number of researchers have investigated how people respond to robot behaviors and how robots should behave so that people can easily understand them (Fong, Nourbakhsh, and Dautenhahn 2003; Breazeal 2004; Kanda et al. 2004). Many insights from developmental or cognitive psychologies have been implemented and examined to see how they affect the human response or whether they help robots produce smooth and natural communication with humans.

Androids are robots whose behavior and appearance are highly anthropomorphized. Developing androids requires contributions from both robotics and cognitive science. To realize a more humanlike android, knowledge from human sciences is also necessary. At the same time, cognitive science researchers can exploit androids for verifying hypotheses in understanding human nature. This new bi-directional, interdisciplinary research framework is called android science (Ishiguro 2005). Under this framework, androids enable us to directly share knowledge between the development of androids in engineering and the understanding of humans in cognitive science.

We have recently developed a 'geminoid,' a new category of robot (figure 1). A geminoid is an android that will

work as a duplicate of an existing person. It appears and behaves as a person and is connected to the person by a computer network. Geminoids have the following capabilities:

- 1) Appearance and behavior highly similar to an existing person. The existence of a real person analogous to the robot enables easy comparison studies. Moreover, if a researcher is used as the original, we can expect that individual to offer meaningful insights into the experiments, which are especially important at the very first stage of a new field of study when beginning from established research methodologies.
- 2) Teleoperation. By introducing manual control, the limitations in current AI technologies can be avoided, enabling long-term, intelligent conversational human-robot interaction experiments. This feature also enables various studies on human characteristics by separating 'body' and 'mind.' In geminoids, the operator (mind) can be easily exchanged, while the robot (body) remains the same.



Figure 1: Geminoid HI-1.

The first geminoid prototype, HI-1, was released in July 2006. Since then, we have encountered several interesting phenomena. Here are some discourses from the geminoid operator.

- When I first saw HI-1 sitting still, it was like looking in a mirror. However, when it began moving, it looked like somebody else, and I couldn't recognize it as myself. This was strange, since we copied my movements to HI-1, and others who know me well say the robot accurately shows my characteristics. This means that we are not objectively recognizing our unconscious movements ourselves.
- While operating HI-1, I find myself unconsciously adapting my movements to the geminoid movements. I felt that my own body is restricted to the movements that HI-1 can make.
- In less than 5 minutes both the visitors and I can quickly adapt to conversation through the geminoid. The visitors recognize and accept the geminoid as me while talking to each other.
- When a visitor pokes HI-1, especially around its face, I get a strong feeling of being poked myself. This is strange, as the system currently provides no tactile feedback.

Before talking through the geminoid, the initial response of the visitors seemingly resembles the reactions seen with previous androids: even though at the very first moment they could not recognize the androids as artificial, they nevertheless soon become nervous while being with the androids. But shortly after having a conversation through the geminoid, they found themselves concentrating on the interaction, and soon the strange feelings vanished. Is intelligence or long-term interaction a crucial factor in overcoming the valley and arriving at an area of natural humanness?

One of our purposes in developing geminoids is to study the nature of human presence. The scientific aspect must answer questions about how humans express and recognize human presence. The technological aspect must realize a teleoperated android that works on behalf of the person remotely accessing it. The following are our current challenges:

- Teleoperation technologies for complex humanlike robots: Methods must be studied to teleoperate the geminoid to convey existence/presence, which is much more complex than traditional teleoperation for industrial robots.
- Psychological test for human existence/presence: We are studying the effect of transmitting *Sonzai-Kan* [the feeling of one's presence] from remote places, for applications such as meeting participation instead of sending the person himself.

Moreover, we are interested in studying existence/presence through cognitive and psychological experiments. For example, we are studying whether the android can represent

the authority of the person himself by comparing the original person and the android.

## References

- Breazeal, C. 2004. Social Interactions in HRI: The Robot View. *IEEE Transactions on Man, Cybernetics and Systems: Part C*, 34: 181–186.
- Fong, T., Nourbakhsh, I., and Dautenhahn, K. 2003. A Survey of Socially Interactive Robots. *Robotics and Autonomous Systems* 42: 143–166.
- Ishiguro, H. 2005. Android Science: Toward a New Cross-Disciplinary Framework. *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, 1–6.
- Kanda, T., Ishiguro, H., Imai, M., and Ono, T. 2004. Development and Evaluation of Interactive Humanoid Robots, *Proceedings of the IEEE*, 1839–1850
- Pascalis, O., Haan, M., and Nelson, C. A. 2002. Is Face Processing Species-Specific During the First Year of Life? *Science*, 296: 1321–1323.